

J3.5 STANDARDIZING THE QUALITY ASSESSMENT OF DATA: PARTNERSHIP ACTIVITIES BETWEEN NCDC AND RCCs

Nathaniel Guttman *
NOAA National Climatic Data Center, Asheville, NC

1. INTRODUCTION

Historically, climatic data have been collected, processed and stored by many different entities in many different ways. Lack of standardization has become a problem when trying to compare analytical results, as well as when serving the interests of the data users. Informally, several climate groups have, over the past several years, discussed the benefits of standardizing processes. An agreement has been reached that: 1) similar kinds of data (e.g., daily data observed at coop and first order) should be treated together with the same rules and algorithms in an integrated manner and 2) algorithms developed by the many entities assessing data should be linked into one unified system so that all basic climate data that are distributed to the public by various agencies are treated in the same manner.

2. PARTNERS

The NCDC supports the long-range goal of standardizing the treatment of all basic data and has begun to actively participate with other groups so that standardization can eventually be achieved. Many data checks are straightforward in that they insure that physical and mathematical rules are followed. Examples are: 1) wind speed must be greater than or equal to zero, 2) a maximum value must be equal or greater than a minimum value, 3) the occurrence of snowfall must be coincident with the occurrence of precipitation. Other straightforward checks involve insuring that the data conform to the observing, coding, and data transmission rules as defined in appropriate manuals and handbooks. Although there is very little disagreement over these types of checks, not all processing centers perform the same checks. Informal discussions among the Regional Climate Centers (RCC) and NCDC are resulting in the collection of routines used by all parties so that a standard, comprehensive suite of checks can be developed.

The NWS has quality control responsibilities for checking data at the local level prior to sending the observations to processing centers such as an RCC or the NCDC. NWS climate services personnel are now collecting information pertaining to the quality control activities at local offices with a goal of standardizing methodologies and operating procedures. The information being collected will be shared with the NCDC, RCCs, and interested State Climatologists.

3. EXAMPLES

There are several data checks that are not straightforward in terms of identifying incorrect data. Two examples are range checks and spatial checks. The first example involves a decision as to whether a value is acceptable if it does not fall within a predetermined range. Some questions concerning this type of decision are: 1) What is an acceptable range? 2) What is an extreme? 3) How are the endpoints of an acceptable range determined? The answers to these questions depend on the points of view and climatological experience of the validators. They also depend on the intended purpose of the quality assurance efforts. Is the purpose to accept a datum if it is "reasonable" or is the purpose to assure correctness? Assimilating the thoughts of all the players into a coordinated, standardized methodology that is accepted by everyone is a difficult process, but a process that is moving forward.

The second example, spatial checks, also involves decisions that reflect the points of view and experience of the validators. These checks compare an observed datum with an estimate that is based on preconceived spatial patterns. Many spatial techniques are available for use, but the decisions resulting from the application of these techniques are not always the same. Method comparisons are essential in gaining an understanding of what some of these different techniques offer. One such comparison was conducted by the NCDC and the High Plains RCC. The NCDC is checking Cooperative Network temperature data by comparing the observation to estimates that are based on gridded NWS first order data. A datum is acceptable if it falls within a broad, empirically determined range. The High Plains RCC has developed a spatial regression technique whereby station comparisons are based on the correlation structure among neighboring stations. The two methods are different in concept, but both are intended to determine if an observed value is reasonable. Both methods were used on the same one year of daily data. Some data were randomly seeded with known errors. For the seeded data, differences in decisions made regarding data acceptability and the magnitude of the error of estimated values were evaluated. Results showed that the NCDC procedure is essentially an extreme value check, and that the spatial regression technique results in fewer decision errors. As a result of the cooperative examination of the two techniques, the spatial regression technique is now being incorporated into the NCDC quality assessment process. But, the partnership has not ended. Once implemented by NCDC, both the High Plains RCC and NCDC will each process the same data independently to insure that the

* Corresponding author address: Nathaniel Guttman, National Climatic Data Center, 151 Patton Avenue, Asheville, NC 28801; e-mail: ned.guttman@noaa.gov.

results of the application of the methodology are identical. Evaluation of the technique will be an ongoing joint effort.

4. SUMMARY

One key to standardizing procedures is communication. A new tracking system, Datzilla, fosters communication among the NCDC, RCCs and NWS. Datzilla allows system users to enter questions and information pertaining to an observed value, estimated value, missing data, processing algorithms, decisions, etc. The system will then track the entry from input to the system to resolution to insure that all issues are handled in an appropriate manner by the responsible parties. All intermediate actions and issues are tracked and viewable by the users.

Standardization is moving forward. Lines of communication have been established, partnerships have been initiated and maintained, workshops have been and will continue to be held, and conference calls are increasing in number. There are many issues that need to be resolved, but all players are actively endorsing the need for cooperation and consensus.